



# Validation of an isotope dilution mass spectrometry (IDMS) measurement procedure for the reliable quantification of steroid hormones in waters

Elodie Mirmont<sup>1</sup> · Amandine Bœuf<sup>1</sup> · Mélissa Charmel<sup>1</sup> · Béatrice Lalère<sup>1</sup> · Sophie Lardy-Fontan<sup>2</sup>

Received: 25 February 2023 / Revised: 6 April 2023 / Accepted: 11 April 2023 / Published online: 29 April 2023  
© Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

Reliable data are compulsory to efficiently monitor pollutants in aquatic environments, particularly steroid hormones that can exert harmful effects at challenging analytical levels below the  $\text{ng L}^{-1}$ . An isotope dilution two-step solid-phase extraction followed by an ultra-performance liquid chromatography separation coupled to tandem mass spectrometry (UPLC-MS/MS) detection method was validated for the quantification of 21 steroid hormones (androgens, estrogens, glucocorticoids, and progestogens) in whole waters. To achieve a realistic and robust assessment of the performances of this method, the validation procedure was conducted using several water samples representative of its intended application. These samples were characterized in terms of concentration of ionic constituents, suspended particulate matter (SPM), and dissolved organic carbon contents (DOC). For estrogens that are part of the European Water Framework Directive Watchlist (17beta-estradiol and estrone), the performances met the European requirements (decision 2015/495/EU) in terms of limit of quantification (LQ) and measurement uncertainty. For 17alpha-ethinylestradiol, the challenging LQ of  $0.035 \text{ ng L}^{-1}$  was reached. More generally, for 15 compounds out of 21, the accuracy, evaluated in intermediate precision conditions at concentrations ranging between  $0.1$  and  $10 \text{ ng L}^{-1}$ , was found to be within a 35% tolerance. The evaluation of the measurement uncertainty was realized following the Guide to the expression of Uncertainty in Measurement. Finally, a water monitoring survey demonstrated the suitability of the method and pointed out the contamination of Belgium rivers by five estrogens (17alpha-ethinylestradiol, estriol, 17alpha-estradiol, 17beta-estradiol, and estrone) and three glucocorticoids (betamethasone, cortisol, and cortisone) which have been up to now poorly documented in European rivers.

**Keywords** Steroids · Ultra-trace quantification · Whole water · Measurement accuracy · Measurement uncertainties

## Introduction

The chronic discharge of Endocrine Disrupting Compounds (EDCs) into the aquatic environment increased consciousness at the political, regulatory, societal, and scientific levels. This is due to the fact that EDCs can pose a threat to biodiversity and exert harmful effects at ultra-trace levels

[1–3]. For example, the exposition to natural and synthetic estrogens has been correlated to the feminization of fish populations [4]. Adverse effects on fish reproduction and modifications of the behavior were also reported following exposition to synthetic steroid hormones (progestogens and glucocorticoids, respectively) [5, 6].

Three estrogens, namely estrone (E1), 17beta-estradiol (17 $\beta$ E2), and 17alpha-ethinylestradiol (17 $\alpha$ EE2), were included in the first European Water Framework Directive (WFD)'s Watchlist in order to generate high-quality data of their occurrence in aquatic environments and contribute in environmental risk assessment [7]. As a consequence, their occurrence was much improved [8], and a ubiquitous state of contamination has been reported worldwide for these estrogens [3, 9, 10] together with other steroid hormones [3, 11–14]. Nonetheless, it should be pointed out that

✉ Amandine Bœuf  
amandine.boeuf@lne.fr

<sup>1</sup> Laboratoire National de métrologie et d'Essai (LNE), 1 rue Gaston Boissier, 75015 Paris, France

<sup>2</sup> Nancy Laboratory for Hydrology, Water Chemistry Unit, ANSES, 40 rue Lionnois, 54 000 Nancy, France

non-estrogenic steroid hormones such as androgens, glucocorticoids, and progestogens have been less investigated to date.

Nowadays, very few multi-class methods can address the challenge of reliable quantification of ultra-trace steroid hormones in waters [15, 16]. Indeed, most of the recently published methods are characterized by inadequate limits of quantification to implement the European WFD [17, 18]. For non-regulated compounds, the challenge resides in quantifying steroid hormones at such low concentration levels, which may pose a risk to environmental and human health. Furthermore, owing to their physicochemical properties, steroid hormones can partition between water and suspended particulate matter (SPM) in water bodies [10, 19, 20]. Even if the European WFD has underlined the importance of analyzing the total fraction of organic micropollutants, including steroid hormones [21], very few monitoring methods can meet this requirement [19, 22]. Nonetheless, this requirement only applies if the European Member States have not empirically demonstrated that the measured concentration of the micro-pollutant in the dissolved fraction is the same as in the total fraction of waters. Indeed, that would prove that targeting the total fraction is irrelevant for environmental monitoring.

Quality of data is crucial since measurement results can underpin key decisions in environmental monitoring. Indeed, reliable data are compulsory for the effectiveness of monitoring, risk assessment, and prioritization, and for the scientific knowledge on steroid hormones. For instance, the cocktail effect and the combined mode of action of steroid hormones need to be further investigated to improve risk assessment [17, 23].

On another note, the comparability of the data at the European level and validation procedures are issues that should be addressed when monitoring chemical substances under the European WFD. In view of harmonizing, a European technical specification, the CEN/TS 16800 describes a framework for the validation of analytical methods in environmental media (solids and water). A similar methodology was given in the French NF T90:210 standard for water quality. However, approaches for estimating performance characteristics such as limit of quantification (LQ) are various (signal-to-noise approach, setting of a maximum allowed tolerance (MAT)... ) and rarely described in sufficient detail in scientific literature. All this contributes to a lack of comparability between the available analytical methods. To assess the performances under operational conditions as closely as possible, matrix effects should be considered in the validation. Moreover, it is of fundamental importance to use as many representative samples as possible to get a realistic and robust assessment of the performances of the method. However, this aspect is generally poorly addressed in published papers. The impact of

matrix effects was scrutinized in another work that highlighted that internal standards could help reduce prejudicial matrix effects in steroid hormones quantification [24]. Three conditions are mandatory to ensure the quality and comparability of measurement results: the establishment of metrological traceability, the demonstration of measurement accuracy, and the estimation of measurement uncertainties. These three conditions have, however, received little attention in the literature.

In this study, a validation approach was developed and optimized for the determination of 21 steroid hormones of four families (androgens, glucocorticoids, estrogens, and progestogens) in whole waters by an isotope dilution two-step solid-phase extraction (SPE) followed by an ultra-performance liquid chromatography separation coupled to tandem mass spectrometry detection (UPLC-MS/MS). Statistical tools described in European standards (CEN/TS 16800 [25] and NF T90:210 [26]) were used to test calibration models and assess the accuracy of the method in representative matrices. Moreover, when possible, certified reference materials (CRMs) traceable to SI units were used to ensure the traceability of measurements. In addition, a thorough evaluation of the measurement uncertainty was performed following the Guide to the expression of Uncertainty in Measurement (GUM) [27]. Finally, to demonstrate its suitability, the method was implemented in a monitoring survey of Belgian surface waters.

## Material and methods

### Chemicals and reagents

Acetonitrile (ACN) and methanol (MeOH) (Baker Analyzed LC-MS grade) were purchased from Atlantic labo (Bruges, France). Water (ULC/MS) and acetone extra dry were acquired from Biosolve chemicals (Dieuze, France). Ethyl acetate (HPLC Plus, purity = 99.90%) and dansyl chloride (HPLC derivatization, purity = 99%) were acquired from Sigma-Aldrich (Saint-Quentin Fallavier, France). A solution of humic acid was prepared following an existing protocol [28], using humic acid sodium salt (technical grade) also provided by Sigma-Aldrich. Formic acid (AnalaR Normapur®, purity > 99%) and EDTA, Na<sub>2</sub> (ethylenediaminetetraacetic acid disodium salt, 0.05 mol L<sup>-1</sup> in aqueous solution) were acquired from VWR (Fontenay-sous-Bois, France). Acetic acid (Optima® LC/MS, purity > 99%) was acquired from Thermo Scientific (Illkirch, France). Evian® water, a commercially available natural mineral water in glass bottles, was used as a reference water. An in-house material containing suspended particulate matter (SPM) was used.

## Selection of native and isotope-labelled analytical standards

21 native compounds belonging to four classes were selected in a previous study [29]:

- 2 androgens (ANDRO): 4-androstenedione (AD) and testosterone (T)
- 6 estrogens (ESTRO): 17alpha-estradiol (17 $\alpha$ E2), 17beta-estradiol (17 $\beta$ E2), 17alpha-ethinylestradiol (17 $\alpha$ EE2), diethylstilbestrol (DES), estriol (E3), and estrone (E1)
- 5 glucocorticoids (GLUCO): betamethasone (BET), cortisol (COL), cortisone (COR), dexamethasone (DEX), and prednisolone (PRED)
- 8 progestogens (PROG): 17alpha-hydroxyprogesterone (17HPT), 21alpha-hydroxyprogesterone (21HPT), chlormadinone acetate (Ac CHLOR), cyproterone acetate (Ac CYP), cyproterone (CYP), drospirenone (DRO), levonorgestrel (LEV), and norethindrone (NOR)

To establish the metrological traceability of measurements, high-purity certified reference materials (CRM) available during the study [30] were selected. CRMs for 17 $\beta$ E2 (CRM 6004-a, purity = 98.4%), COL (CRM 6007-a, purity = 99.3%), and T (CRM 6002-a, purity = 99.8%) were provided by the National Metrology Institute of Japan (NMIJ). CRM for 17HPT (S041, purity = 98.7%) was provided by the National Measurement Institute of Australia (NMIA).

For other native compounds, analytical standards with the highest purity were selected.

The choice of isotope-labelled compounds was a crucial step. Labelled compounds from a sufficient mass difference with target compounds to prevent a detrimental impact on quantification should be selected. For example, quantification biases were reported when using cortisol-d<sub>2</sub>: an interference was reported with a natural isotope of cortisol present in the sample. Similar issues were reported with the use of testosterone-d<sub>2</sub> [31]. Furthermore, H/D exchanges between labelled compounds and reconstitution solvents have been reported by Davison et al. with increased exchanges at low pH values [32]. Labelled compounds with <sup>13</sup>C are generally recognized as more reliable, owing to their position within the carbon backbone of steroid hormones, making them unavailable for any exchange. All these considerations led to systematically rejecting compounds with less than 3 Da of difference with their native analogous and, when possible, to favor compounds labelled with <sup>13</sup>C. Twelve isotope-labelled compounds were finally chosen, considering their labelling quality, physicochemical properties, availability, and cost. Isotope dilution was implemented for every analyte using labelled analogue (deuterated or <sup>13</sup>C) except for 17 $\alpha$ E2, E3, BET, COR, COL, 21HPT, Ac CHLOR, Ac CYP, and CYP.

All the native and isotope-labelled compounds (<sup>13</sup>C and <sup>2</sup>H) were purchased as pure analytical standards or solutions (in methanol or dioxane at 0.1 mg mL<sup>-1</sup>) from LGC standards (Molsheim, France), Merck KGaA (Darmstadt, Germany), Toronto Research Chemicals Inc. (Ontario, Canada), Cambridge Isotope Laboratories Inc. (Tewksbury, USA), and Cayman Chemical (Ann Arbor, USA).

CAS numbers and purity of analytical standards, as well as some physicochemical properties (log K<sub>ow</sub> and pKa), are available in Supplementary Information for the 21 native compounds (Table S1) and the 12 isotope-labelled compounds (Table S2).

## Preparation of stock and working solutions

Individual stock solutions were gravimetrically prepared at about 0.1 mg mL<sup>-1</sup> in methanol for all the native and isotope-labelled compounds and were analyzed to check for cross-contamination (absence of another selected compound).

Working solutions of the target compounds and the isotope-labelled compounds were gravimetrically prepared by diluting the individual stock solutions in methanol.

Specifically, for the validation of the method, two independent batches of individual solutions of the native compounds were prepared when possible: the first for the preparation of calibration samples and the second for the spiking of samples.

All stock and working solutions were stored in the dark at  $-20 \pm 5$  °C. No instability of compounds in working solutions was detected during a storage period of 5 months.

## Measurement procedure

The analytical procedure used in this study was thoroughly described by Mirmont et al. [29]. Each critical step of the measurement procedure was controlled by weighing.

To summarize, whole water samples (1 L) were spiked with EDTA (0.1%; v/v) and isotope-labelled compounds at concentrations ranging between 0.5 and 2.5 ng L<sup>-1</sup> for 17 $\alpha$ EE2-d<sub>4</sub> and LEV-d<sub>6</sub>, respectively. After extraction on C18 Atlantic® Ready Disk (Biotage, Uppsala, Sweden) using a Horizon Technology SPEDEX®-4790, purification of samples using Supelclean™ LC-NH<sub>2</sub> SPE (500 mg, 6 mL) cartridge (Merck, Darmstadt, Germany) was performed.

As two chromatographic runs were required for the quantification of trace compounds in whole water, each extract was divided into two aliquots: one aliquot was used for the analysis of GLUCO in the negative mode, and the analysis of ANDRO and PROG in the positive mode; the other aliquot was dedicated to the analysis of ESTRO in the positive mode after a dansylation step.

Liquid chromatography was performed using an Acquity® UPLC H-Class system (Waters, Guyancourt, France). Separation was achieved using an Accucore™ Biphenyl column (2.6  $\mu\text{m}$ ; 2.1  $\times$  100 mm) equipped with an Accucore™ Biphenyl pre-column (2.6  $\mu\text{m}$ ; 2.1  $\times$  10 mm) and a pre-filter (0.2  $\mu\text{m}$ ) (Thermo Scientific, Waltham, USA).

The UPLC system was coupled to a Xevo TQ-MS® triple quadrupole mass spectrometer (Waters, Guyancourt, France) equipped with an electrospray ionization (ESI) source. The mass spectrometer was set with a capillary voltage of 3 kV in positive mode and  $-2$  kV in negative mode. Source and desolvation temperatures were set at 150  $^{\circ}\text{C}$  and 650  $^{\circ}\text{C}$ , respectively. Desolvation and cone gas were set at 1000 and 50  $\text{L h}^{-1}$ , respectively. The MS acquisition was performed in Multiple Reaction Monitoring (MRM) mode, and data were processed with TargetLynx™ (Waters).

For each compound, one transition was selected for the quantification, and one transition was selected for the confirmation. Both ions and MS parameters for the 21 target compounds and the 12 isotope-labelled compounds are detailed in Mirmont et al. [29].

## Quality controls

To check the initial system performances and to monitor any prejudicial loss of sensitivity during the analytical runs, mixtures of target compounds (1  $\text{ng mL}^{-1}$  and 10  $\text{ng mL}^{-1}$ ) and isotope-labelled compounds (25  $\text{ng mL}^{-1}$ ) were injected regularly in each analytical run. A 20% tolerance was stated based on previous laboratory experience with chromatographic peak areas. Moreover, possible carry-over due to memory effects from one sample to another during analytical runs was scrutinized by the injection of solvent blanks (same conditions as samples) after every sample.

Besides, to check any contamination during the overall analytical workflow, blanks constituted of Evian® water spiked with isotope-labelled compounds at a concentration ranging between 0.5 and 2.5  $\text{ng L}^{-1}$  for 17 $\alpha$ EE2-d4 and LEV-d6, respectively, were used in each analytical series.

Solvent and procedural blanks were systematically looked into in order to check for the absence of compounds of interest or interferences.

Positive controls were examined in each analytical run. They were constituted of Evian® water samples spiked with the compounds at concentrations ranging from 0.1 for 17 $\alpha$ EE2 to 5.0  $\text{ng L}^{-1}$  for LEV and with the isotope-labelled compounds at concentrations ranging from 0.5 for 17 $\alpha$ EE2-d4 to 2.5  $\text{ng L}^{-1}$  for LEV-d6 and NOR-d6.

Confirmation of identification of target compounds was performed, fulfilling the ISO 21253-1:2019 requirements: retention time with a tolerance of 2.5%, monitoring of two distinct transitions and their abundance ratio (based on peak

area) with a 30% tolerance between samples and calibration samples [33].

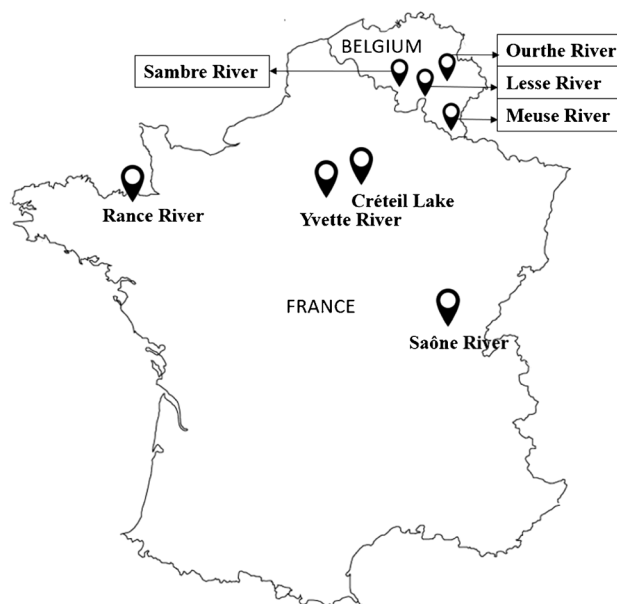
## Method validation

Method validation was performed using some of the specific statistical tools provided in the NF T90:210 standard.

## Selection of representative samples

Seven water samples with an extensive range of physicochemical properties were selected as representative of environmental conditions to which the method will be applied for monitoring in surface waters. Water samples were collected in 4 different locations in France (Fig. 1): the Rance, the Saône, the Yvette Rivers, and the Créteil Lake. Furthermore, two synthetic samples exempt from all target compounds were used as blanks to assess the accuracy at the LQ level. To mimic environmental waters, these two synthetic samples were prepared by spiking Evian® mineral water with humic acid to a dissolved organic carbon (DOC) concentration of 5  $\text{mg L}^{-1}$  or with suspended particulate matter (SPM) at a concentration of 50  $\text{mg L}^{-1}$ . These levels were derived from concentrations of DOC and SPM found in average French surface waters [34].

The physicochemical properties of the seven water samples used for method validation are given in Supplementary Information (Table S3).



**Fig. 1** Location of the selected sites for method validation (in France) and for environmental monitoring (in Belgium, text in boxes)



## Calibration model

The quantity  $Q$  of a target compound in a sample can be calculated thanks to a calibration function described in the following equation (Eq. 1):

$$Q = \left(a \frac{A}{A^*} + b\right) Q^* \quad (1)$$

where  $Q^*$  is the quantity of the internal standard (isotope-labelled compound);  $\frac{A}{A^*}$  is the chromatographic peak area ratio between the target compound and its internal standard; and  $a$  and  $b$  are the slope and the  $y$ -intercept of the calibration function.

Calibration samples with the closest area ratios  $\frac{A}{A^*}$  observed in samples were chosen to custom-build a suitable calibration curve for each target compound. According to the target compounds and linearity of the calibration model, one or two calibration models were considered for the quantification. Multipoint calibration curves consisted of 5 to 11 points, and for each compound, the previous equation was determined from the calibration model. Calibration samples were gravimetrically prepared in a mixture of water and acetonitrile (same conditions as for sample analyses). The concentration of target compounds ranged from 0.1 to 62.5 ng mL<sup>-1</sup>, and the concentration of isotope-labelled compounds ranged from 2.5 to 12.5 ng mL<sup>-1</sup>. Concentrations of the different compounds in calibration samples are given in Supplementary Information (Table S4).

Calibration models were evaluated in intermediate precision conditions over seven different days, with calibration samples randomly injected in triplicate under repeatability conditions.

Calibration criteria were set as follows: the back-calculated concentrations of calibration samples should be within  $\pm 20\%$  of the nominal value at the lowest concentrated calibration sample, and within  $\pm 15\%$  for the other calibration samples to be considered acceptable.

## Establishment of measurement traceability to the SI units

Solutions prepared using high-purity CRM for COL at 30.0 ng mL<sup>-1</sup>, for 17HPT at 3.0 ng mL<sup>-1</sup>, for T at 1.5 ng mL<sup>-1</sup>, and for 17 $\beta$ E2 at 0.3 ng mL<sup>-1</sup> were analyzed to assess the calibration trueness and establish metrological traceability of measurements to SI units. Bias was calculated for every solution in repeatability conditions. Bias was defined in the following equation (Eq. 2) as:

$$\text{Bias (\%)} = \frac{C_{\text{measured}} - C_{\text{theoretical}}}{C_{\text{theoretical}}} \times 100 \quad (2)$$

where  $C_{\text{measured}}$  is the mean measured concentration in solutions ( $n=3$ ) and  $C_{\text{theoretical}}$  is the theoretical concentration in solutions determined gravimetrically.

## Method accuracy

According to the International Vocabulary of Metrology (VIM), measurement accuracy is defined as the “closeness of agreement between a measured quantity value and a true quantity value of a measurand” [35]. To assess the accuracy of measurement results, the first step is to estimate the measurement precision and bias. The second step consists of setting a maximum allowed tolerance (MAT). For that purpose, the two following inequalities have to be verified (Eq. 3 and Eq. 4):

$$\text{MAT(\%)} < B(\%) - \frac{2S_{\text{FI}}}{C_{\text{theoretical}}} \quad (3)$$

$$B(\%) + \frac{2S_{\text{FI}}}{C_{\text{theoretical}}} < \text{MAT(\%)} \quad (4)$$

where MAT is the maximum allowed tolerance,  $B$  is the bias as described in the “Establishment of measurement traceability to the SI units” section, and  $S_{\text{FI}}$  is the standard deviation in intermediate precision conditions.

The MAT was fixed depending on the compound and the level of concentration. As requested by the NF T90-210 standard, a MAT was set at 60% at the LQ and 35% at concentrations above the LQ.

As no matrix-matched CRMs were available, the pure analytical standards purchased were considered as reference and working solutions were used to gravimetrically spike samples from the seven selected matrices (see “Selection of representative samples” section). Independent duplicate samples (1 L) of each matrix were prepared according to the procedure described in the “Measurement procedure” section and analyzed in intermediate precision conditions for each investigated level of concentration (one operator, seven different days (1 day/matrix), preparation of two replicates in repeatability conditions for each level of concentration).

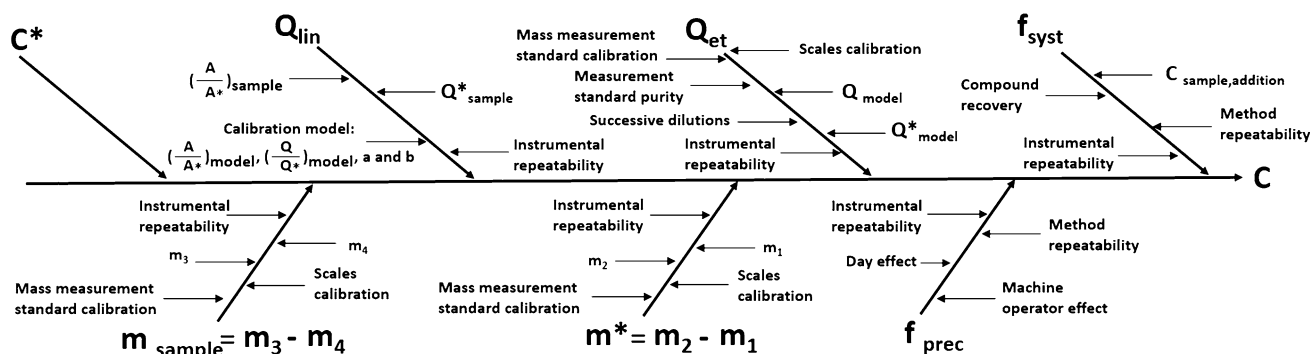
## Measurement uncertainty

Every measurement result should be expressed with its associated expanded uncertainty  $U$  to allow the comparison of measurement results. A coverage factor ( $k$ ) of 2 is generally chosen with a confidence level of 95%.

Measurement uncertainties were evaluated by following the Guide to the expression of Uncertainty in Measurement (GUM).

As a first step, an Ishikawa diagram, given in Fig. 2, was constructed by listing all the parameters and potential sources of errors that may influence the calculation of the concentration of a target compound in a sample. This diagram is a precious tool in the evaluation of the global uncertainty budget.

Then, the concentration of a target compound in a sample was expressed in a mathematical form with the following equation (Eq. 5):



**Fig. 2** Ishikawa diagram.  $C$ , concentration of the target compound in the sample;  $C^*$ , concentration of the internal standard in the spiking solution;  $Q_{lin}$ , modeled quantity ratio  $\frac{Q}{Q^*}$  in the sample calculated thanks to constants  $a$  and  $b$  from the linear calibration model;  $\frac{A}{A^*}$ , area ratio in the sample;  $Q_{et}$ , corrective factor linked to the uncertainty on concentration of calibration samples;  $f_{syst}$ , correction factor linked to

the variability of the bias (systematic measurement error);  $m_3$ , mass of the empty sample bottle;  $m_4$ , mass of the bottle after sampling;  $m_2$ , mass of the empty bottle before addition of internal standards;  $m_1$ , mass of the bottle after the addition of internal standards;  $f_{prec}$ , correction factor linked to the intermediate precision of the method (random measurement error)

$$C = \frac{C^*(Q_{lin})(Q_{et})(m_2 - m_1)}{(m_3 - m_4)/1000} + f_{prec} + f_{syst} \quad (5)$$

where  $C$  is the concentration of the target compound in the sample in  $\text{ng L}^{-1}$ ,  $C^*$  is the concentration of the internal standard in the spiking solution in  $\text{ng g}^{-1}$ ,  $Q_{lin}$  is the quantity ratio  $\frac{Q}{Q^*}$  in the sample calculated using the constants  $a$  and  $b$  from the linear calibration model,  $Q_{et}$  is a corrective factor linked to the uncertainty linked to the preparation of calibration samples,  $m_2$  is the mass of the empty bottle before addition of internal standards in g,  $m_1$  is the mass of the bottle after the addition of internal standards in g,  $m_3$  is the mass of the empty bottle before sampling in g,  $m_4$  is the mass of the bottle after sampling in g, 1000 is a conversion factor from  $\text{ng g}^{-1}$  to  $\text{ng L}^{-1}$  considering that sample density is equal to water's density,  $f_{prec}$  is a correction factor linked to the uncertainty associated to the intermediate precision of the method (random measurement error), and  $f_{syst}$  is a correction factor linked to the uncertainty associated to the variability of the bias (systematic measurement error).

Corrective factors  $Q_{et}$ ,  $f_{prec}$ , and  $f_{syst}$  are respectively equal to 1, 0, and 0. They do not participate in target compound concentration calculation. However, they are described in the mathematical model given in Eq. 5 because their associated relative uncertainties contribute to the global uncertainty budget.

Standard measurement uncertainties of each input value in the quantification model of Eq. 5 were evaluated.  $u(Q_{lin})$  was calculated by polynomial regression using the least square method.  $u(Q_{et})$  was calculated considering uncertainty on standard purity (given by the supplier's certificates) and uncertainty of weighing due to preparations and dilutions.  $u(m_i)$  related to all weighted masses were given by the latest calibration certificate of each scale used. The uncertainty contribution

of the working solution of isotope-labelled compounds was not included since the same solution was used for the preparation of the calibration samples and spiked samples.  $u(f_{prec})$  is the standard deviation of the independently measured concentrations of the same theoretical concentration in intermediate precision conditions (one operator, seven different days (1 day/matrix investigated), preparation of two replicates in repeatability conditions for each level of concentration).  $u(f_{syst})$  is the standard deviation of independent recoveries under intermediate precision conditions of the same known quantity of a given compound added in the sample with (Eq. 6):

$$u(f_{syst}) = \sqrt{b_{rms}^2 + u_{add}^2} \quad (6)$$

where  $b_{rms}$  is the root mean square of the deviation from a 100% recovery (%),  $u_{add}$  is the uncertainty in the concentration of the added compound (%), and  $u_{add}$  (%) is obtained from the following equation (Eq. 7):

$$u_{add} = \sqrt{u_v^2 + u_{conc}^2} \quad (7)$$

where  $u_v$  is the relative uncertainty component of the volume added (with a gravimetric control) (%) and  $u_{conc}$  is the relative uncertainty component of the concentration of the spiking solution (%)

Finally, the combined standard uncertainty was calculated following the general law of propagation of uncertainty using Wincert® software (version 3.13.0311.0026).

## Application to environmental monitoring

A sampling campaign was conducted in the Walloon part of the Meuse District in Belgium at four locations in July and December 2020: the Meuse, the Ourthe, the Lesse, and

the Sambre rivers (Fig. 1). These monitoring stations were selected for their representativeness of different pressures: livestock farming and urban pressure.

Environmental grab samples were collected in amber glass bottles previously calcinated at 450 °C for 4 h. Samples were transported and stored at  $4 \pm 3$  °C. Sample preparation (see the “[Measurement procedure](#)” section) was performed within 24 to 48 h after sampling.

## Results and discussion

### Method validation

#### Calibration model

The bias from the back-calculated concentrations of calibration samples was considered satisfying for 16 compounds out of 21 for the seven analytical series performed under intermediate precision conditions. For five compounds, a higher bias was observed. For E3, differences of 22 and 25% were observed at the lowest concentration calibration sample (E3 concentration of 0.4 ng mL<sup>-1</sup>). Similar observations were found for 21HPT, COR, and LEV with a bias of 22, 24, and 30%, respectively, at the lowest concentration calibration sample (concentration of 0.3, 1.0, and 3.0 ng mL<sup>-1</sup>, respectively). Lastly, for DES, a difference of 25% was observed at a concentration of 5.0 ng mL<sup>-1</sup>.

These biases may originate from the lack of isotope-labelled compounds for E3, COR, and 21HPT. The quality of labelling may also contribute to these biases.

#### Establishment of measurement traceability to the SI units

A bias of less than 5% was obtained when standard solutions of high-purity CRM (17βE2, 17HPT, COL, and T) were used, demonstrating the trueness of the calibration. The detailed data are summarized in Table 1. Considering the state of the art on available technical resources, this result was considered rather satisfying even if the metrological traceability for every target compound shall, as soon as possible, be established to guarantee the quality and comparability of data.

#### Method accuracy

The accuracy profiles, e.g., a graphical representation of accuracy with concentration, starting at the LQ, are shown in Fig. 3 for the 21 target compounds of this study.

**Estrogens** Validated LQs ranged between 0.035 ng L<sup>-1</sup> for 17αEE2 and 0.5 ng L<sup>-1</sup> for DES. For 17βE2, E1, 17αE2, and E3, a LQ of 0.1 ng L<sup>-1</sup> was validated. The confirmation of

**Table 1** Analysis of high-purity CRM standard solutions

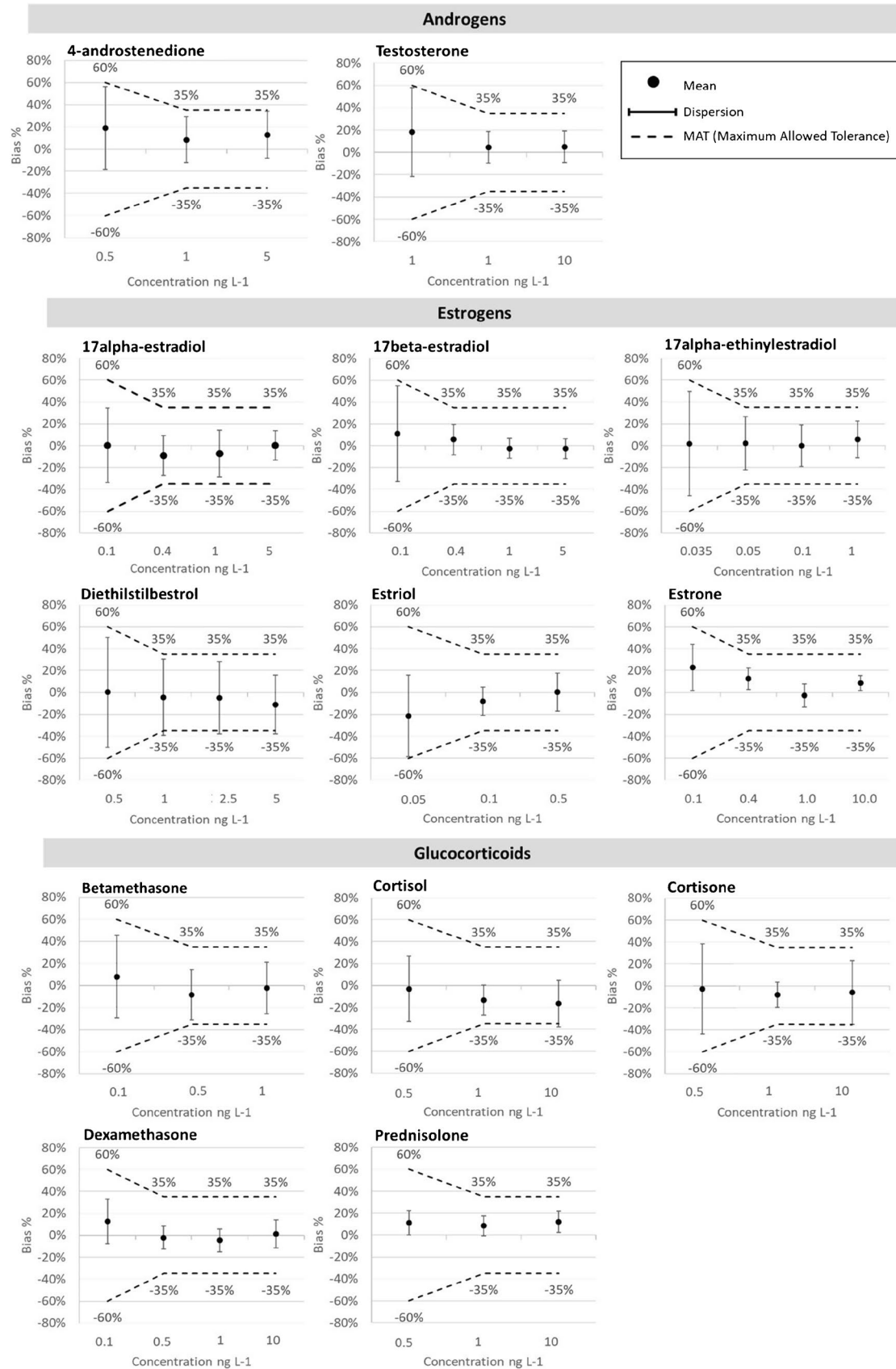
Compounds	CRM reference	Concentration of solution in ng mL <sup>-1</sup>	Bias (%)
COL	NMIJ CRM 6007-a	30	4.0%
17HPT	NMIA S041	3.0	-0.2%
T	NMIJ CRM 6002-a	1.5	4.0%
17βE2	NMIJ CRM 6004-a	0.3	4.0%

the identification according to the four criteria detailed in the “[Quality controls](#)” section was achieved for all estrogens at the LQ level.

Nonetheless, all target estrogens, except DES, were quantified or detected in non-spiked samples from the Saône, the Yvette Rivers, and the Créteil Lake. Data were not considered when natural levels were close to or higher than a specific spiking level. These observations highlight one of the struggles in the validation of analytical methods using real matrices: the difficulty of finding matrices exempt from target compounds. This is particularly problematic for the estimation and validation of analytical performances at trace levels, as it is the case in this study. Indeed, it appears rather impossible to assess the LQ when natural occurrence levels are higher than or close to the LQ. Finally, for 17βE2, the LQ of the method is adequate, considering the WFD Average Annual Environmental Quality Standards (AA-EQS) in surface waters lowered at 0.18 ng L<sup>-1</sup> in November 2021 [36]. For 17αEE2, the AA-EQS was reduced to 0.017 ng L<sup>-1</sup> in November 2021, but the still challenging level of 0.035 ng L<sup>-1</sup> was reached. For E1, the 3.6 ng L<sup>-1</sup> Predicted No-Effect Concentration (PNEC) is approximately one hundred times higher than the validated LQ, demonstrating it is compatible with the implementation of the method for the characterization of the aquatic environment. Glineur et al. and Lardy-Fontan et al. have developed methods for the quantification of traces of estrogens in the total fraction of raw surface waters. Lardy-Fontan et al., with an approach similar to the one used in this study, reported higher LQ of 0.4 ng L<sup>-1</sup> for E1 and 17βE2 and a higher LQ of 0.1 ng L<sup>-1</sup> for 17αEE2 [10]. Glineur et al. reported lower LQs of 0.021, 0.053, 0.028, and 0.036 ng L<sup>-1</sup> respectively for 17αEE2, E1, 17βE2, and E3; however, they used a signal-to-noise approach to estimate the LQ and only estrogens were targeted in their method [37].

As shown in Fig. 3, the accuracy of the present method was found to be within 35% tolerance at the higher concentration levels for the three WFD-regulated estrogens 17αEE2, 17βE2, and E1, as well as for other estrogens (17αE2 and E3). For DES, the criteria of 35% tolerance could not be met and had to be extended to a 40% tolerance.

Lardy-Fontan et al., with a similar approach to this study, demonstrated using Evian® mineral water and Oise river the



**Fig. 3** Accuracy at different levels of concentrations for the 21 target compounds of this study



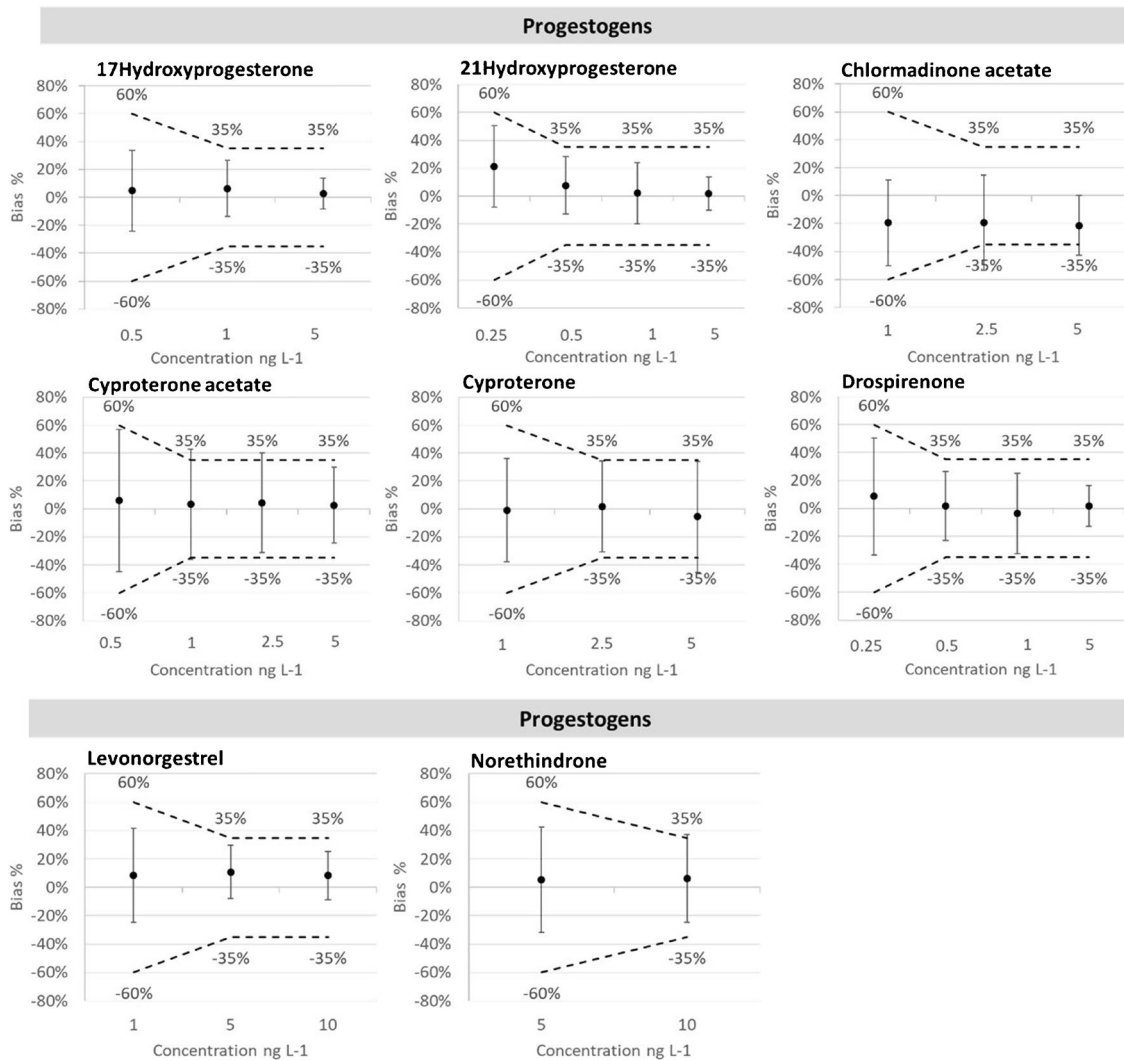


Fig. 3 (continued)

accuracy of the method with a 30% tolerance for E1 and 17 $\beta$ E2 at 1.2 ng L<sup>-1</sup> and 3.6 ng L<sup>-1</sup>, respectively and with a 40% tolerance for 17 $\alpha$ EE2 at 0.4 ng L<sup>-1</sup> [10]. Glineur et al. fixed a tolerance of 40% and applied the same approach to analyze estrogens in the total fraction of seven natural surface waters. The accuracy was finally demonstrated by Glineur et al. at 0.5 ng L<sup>-1</sup> and 2.0 ng L<sup>-1</sup> for 17 $\beta$ E2, 17 $\alpha$ EE2, and E3 and at 2.0 ng L<sup>-1</sup> and 8.0 ng L<sup>-1</sup> for E1 [37]. Nonetheless, these two studies only targeted estrogens, contrary to the present work in which four families of steroid hormones are targeted.

Consequently, the present study's results were considered satisfactory considering data on the literature, WFD regulation requirements, and knowledge needed for DES, for which environmental occurrence is very few documented to date.

**Androgens** Validated LQs were 0.5 and 1.0 ng L<sup>-1</sup> for T and AD, respectively. At these levels of concentration, the

confirmation of the identification of T and AD was possible according to the criteria described in the “Quality controls” section. These LQs are relevant for environmental monitoring considering the 100 and 14 ng L<sup>-1</sup> PNEC found in the literature for T and AD, respectively [38].

As shown in Fig. 3, the accuracy of the method was demonstrated with a 35% tolerance at higher levels of concentration for AD and T. To the author's knowledge, there is no other study in which androgens were investigated in whole water samples.

**Progesterogens** As shown in Fig. 3, validated LQs ranged between 0.25 ng L<sup>-1</sup> for 21HPT and DRO and 5.0 ng L<sup>-1</sup> for NOR. For 17HPT, a LQ of 0.5 ng L<sup>-1</sup> was validated. For Ac CYP, CYP, and LEV, a LQ of 1.0 ng L<sup>-1</sup> was reached. For Ac CHLOR, a LQ of 2.5 ng L<sup>-1</sup> was validated. The confirmation of the identification of all progesterogens was

possible at the LQ level. For DRO, the validated LQ is relevant considering the  $2.0 \text{ ng L}^{-1}$  PNEC [38]. However, for the other target compounds, to date, no PNEC are available, and complementary information on ecotoxicity is still needed to confirm the relevance of these LQs for environmental monitoring.

Shen et al. developed a method for the simultaneous quantification of 61 progestogens in filtrated river water, and reported LQs ranging between  $0.03$  and  $0.4 \text{ ng L}^{-1}$  with a signal-to-noise approach [39]. These LQs are lower than in the present work. Nonetheless, it should be pointed out that this method only targeted one class of steroid hormones in filtrated samples. To the author's knowledge, there is no other study in which progestogens were investigated in whole water samples.

For 17HPT, 21HPT, LEV, and DRO, the accuracy of the present method met the 35% tolerance at high concentration levels. However, if a tolerance of 40% is chosen, the accuracy of the method was verified for NOR but not for Ac CYP, Ac CHLOR, and CYP. The present method gives semi-quantitative information on the occurrence of these few documented compounds. Globally, for these compounds, the performances of the method could be improved by implementing isotope dilution, which is a powerful tool for compensating eventual losses during sample preparation and/or alteration of the signal during sample analysis [24].

**Glucocorticoids** Validated LQs ranged between  $0.1 \text{ ng L}^{-1}$  for BET and DEX and  $0.5 \text{ ng L}^{-1}$  for COL, COR, and PRED. Confirmation of the identification was impossible for GLUCO because no signal was detected for the confirmation transition, which is about ten times lower than the quantification one. However, the confirmation of the identification of GLUCO was possible at concentrations from  $10 \text{ ng L}^{-1}$ .

Validated LQs were found to be relevant for PRED and COL, considering their respective 230 and  $2000 \text{ ng L}^{-1}$  PNEC [38]. However, more information is needed for other glucocorticoids.

With a signal-to-noise approach, Shen et al. reported lower LQs ranging between  $0.01$  and  $0.13 \text{ ng L}^{-1}$  for the determination of 68 glucocorticoids in filtrated river water but no other class of steroid hormones were investigated in this study as in present work [40]. To the author's knowledge, there is no other study in which glucocorticoids were investigated in whole water samples.

For BET, COR, DEX, and PRED, the accuracy of the present method was demonstrated with a 35% tolerance. Considering a higher tolerance of 40%, the accuracy of the method was verified for COL. The performances of the method were considered satisfactory. Particularly, for COL, performances could be improved by implementing isotope dilution.

In the literature, only a few multi-class and ultra-trace methods were found to give a reliable quantification of steroid hormones in river waters. A method targeting all four investigated classes of compounds, reaching the sub  $\text{ng L}^{-1}$  range in the dissolved fraction of surface waters for T, COL, COR, and NOR with lower LQ at approximately  $0.3 \text{ ng L}^{-1}$  and higher levels for all the other compounds investigated, was documented [41]. With a signal-to-noise approach in filtrated surface water, Goeury et al. reported a LQ of  $0.30 \text{ ng L}^{-1}$  for E1,  $17\alpha\text{E}2$ ,  $17\beta\text{E}2$ , and E3; of  $1.50 \text{ ng L}^{-1}$  for  $17\alpha\text{EE}2$ , LEV, and NOR; and of  $0.45 \text{ ng L}^{-1}$  for AD and T [42]. It should nonetheless be pointed out that comparison of method LQ is a rather complex task since their evaluation has been rarely described in detail (in terms of the matrix used and the statistical approach chosen for data treatment). Moreover, the LQ is often only estimated considering the dissolved fraction of compounds. Furthermore, to the author's knowledge there is no other study in which the estimation and validation of LQ were conducted in matrices with such contrasted physicochemical properties as in present work. Nevertheless, this methodology allows a more robust assessment of method performances. Indeed, Tavazzi et al. chose Milli-Q water [43], and other authors such as Gong et al. or Goeury et al. chose a single natural matrix [13, 43 for the validation of LQ. The same observations were drawn for the accuracy of analytical methods. These discrepancies in methodology strengthen once again the difficulty of comparing method performances.

## Measurement uncertainty

Expanded measurement uncertainties ( $k=2$ ) estimated with the GUM approach are given in Table 2 for each target compound. These values were obtained under operational conditions in intermediate precision conditions (one operator, seven different days (1 day/matrix investigated), preparation of two replicates in repeatability conditions for each concentration level).

As shown in Table 2, at the LQ level, uncertainties ranged between 58 and 75% for WFD-regulated estrogens. At  $1 \text{ ng L}^{-1}$ , uncertainties were reduced to values ranging between 14 and 25% for these compounds. Similar observations were found for  $17\alpha\text{E}2$  and E3. For DES, uncertainties were higher than those for other estrogens, with a value of 45% at  $5.0 \text{ ng L}^{-1}$ . This is consistent with the performances of the method in terms of accuracy, detailed in the “Method accuracy” section. Moreover, the estimated uncertainty is, for E1, in line with the QAQC (2009/90/CE) [44] requirements that set a maximal acceptable uncertainty of 50% at its AA-EQS (PNEC of  $3.6 \text{ ng L}^{-1}$ ). Concerning  $17\alpha\text{EE}2$  and  $17\beta\text{E}2$ , this prerequisite was not met considering their recently updated AA-EQS in November 2021 (lowered at  $0.017$  and  $0.18 \text{ ng L}^{-1}$  respectively).

**Table 2** Expanded measurement uncertainty ( $k=2$ ) for 21 target compounds

Compound	C (ng L <sup>-1</sup> )	U% ( $k=2$ ) GUM
<b>Estrogens</b>		
17 $\alpha$ EE2	<b>0.035</b>	<b>75%</b>
	0.05	36%
	0.1	26%
	1.0	25%
17 $\beta$ E2	<b>0.1</b>	<b>71%</b>
	0.4	23%
	1.0	14%
	5.0	14%
E1	<b>0.1</b>	<b>58%</b>
	0.4	29%
	1.0	22%
	10.0	20%
17 $\alpha$ E2	<b>0.1</b>	<b>51%</b>
	5.0	20%
E3	<b>0.1</b>	<b>69%</b>
	5.0	25%
DES	<b>0.5</b>	<b>60%</b>
	5.0	45%
<b>Progestogens</b>		
Ac CHLOR	<b>2.5</b>	<b>64%</b>
	5.0	55%
Ac CYP	<b>1.0</b>	<b>53%</b>
	5.0	37%
CYP	<b>1.0</b>	<b>53%</b>
	5.0	57%
LEV	<b>1.0</b>	<b>49%</b>
	10.0	30%
DRO	<b>0.25</b>	<b>62%</b>
	5.0	21%
17HPT	<b>0.5</b>	<b>45%</b>
	5.0	16%
21HPT	<b>0.25</b>	<b>59%</b>
	5.0	18%
NOR	<b>5.0</b>	<b>51%</b>
	10.0	43%
<b>Glucocorticoids</b>		
BET	<b>0.1</b>	<b>56%</b>
	1.0	39%
DEX	<b>0.1</b>	<b>44%</b>
	10.0	18%
COL	<b>0.5</b>	<b>43%</b>
	10.0	46%
COR	<b>0.5</b>	<b>57%</b>
	10.0	43%
PRED	<b>0.5</b>	<b>27%</b>
	10.0	28%

**Table 2** (continued)

Compound	C (ng L <sup>-1</sup> )	U% ( $k=2$ ) GUM
<b>Androgens</b>		
T	<b>0.5</b>	<b>61%</b>
	10.0	21%
AD	<b>1.0</b>	<b>28%</b>
	5.0	37%

For WFD-regulated estrogens, data are given for all investigated levels of concentration. For other compounds, data are, for reasons of clarity, given at the LQ level and at the higher investigated level of concentration

Values in bold correspond to the LQ level

Nonetheless, for 17 $\alpha$ EE2 at the challenging concentration of 0.05 ng L<sup>-1</sup>, uncertainty was only 36%. For 17 $\beta$ E2, uncertainties are only at 23% at 0.4 ng L<sup>-1</sup>. Other steroid hormones showed the same trends, with uncertainties ranging between 27 and 66% at the LQ and between 16 and 57% at higher concentrations. Uncertainties were globally higher for compounds for which accuracy was not verified with a 35% tolerance (Ac CHLOR, Ac CYP, CYP, NOR, and COL).

In most cases, the main contributions to the global uncertainty were the intermediate precision and variability of the bias. However, for some compounds (17 $\alpha$ EE2 and Ac CHLOR, for example), the calibration model was also found to contribute to the global uncertainty (see Supplemental Information (Fig. S1)). These results highlighted the advantage of using the GUM approach as it allows a close evaluation of measurement uncertainties thanks to mathematical modelling compared to other approaches like the ISO 11352 one, where only the main sources of uncertainty are considered (empirical approach) [45]. Considering the targeted concentration levels and the variety of target compounds investigated, uncertainties were considered satisfactory.

To the authors' knowledge, there is no method by which uncertainties can be compared with regard to the investigated compounds in comparable conditions. Tavazzi et al. described a method for the characterization of the three WFD-regulated estrogens in the total fraction of surface waters. Using the GUM approach, expanded uncertainties ( $k=2$ ) of 12% for 17 $\alpha$ EE2 at 0.05 ng L<sup>-1</sup> and 6 and 16% at 0.3 ng L<sup>-1</sup> for 17 $\beta$ E2 and E1, respectively, were reported. However, these very low expanded uncertainties were estimated under intermediate precision conditions using ultrapure water (Milli-Q®) [43] and not natural waters, as it was the case in this present study). Lardy-Fontan et al. used the top-down approach described in the ISO 11352:2012 to assess the measurement uncertainty for the analysis of these three compounds in the total fraction of surface waters (Oise River with SPM > 100 mg L<sup>-1</sup> and DOC > 5 mg L<sup>-1</sup>).

Reported expanded uncertainty ( $k=2$ ) was 50% for  $17\alpha\text{EE}2$  at  $0.1 \text{ ng L}^{-1}$  and 35% for  $17\beta\text{E}2$  and  $\text{E}1$  at  $0.4 \text{ ng L}^{-1}$  [10].

Considering the few available data in the literature and WFD regulation requirements, the present method showed performances in terms of LQ, accuracy, and uncertainty in real matrices that were considered satisfactory and sufficient for its intended use. Therefore, the developed method offers the potential of becoming a candidate reference method since it shows the best possible performances in the light of the state of the art of available technical resources. Indeed, the method aims at bringing reliable, comparable, and traceable data. Finally, the developed method was implemented in a monitoring survey.

### Fitness for purpose for environmental monitoring

Table 3 presents an overview of measured concentrations in samples collected during two campaigns in July and December 2020 on Belgian rivers: the Meuse, the Ourthe, the Lesse, and the Sambre Rivers (Fig. 1). Chromatograms of the detected compounds in the Sambre River during the campaign of December 2020 are available in Supplementary Information (Fig. S2).

Only five estrogens ( $17\alpha\text{EE}2$ ,  $\text{E}3$ ,  $17\alpha\text{E}2$ ,  $17\beta\text{E}2$ , and  $\text{E}1$ ) and three glucocorticoids (BET, COL, and COR) were detected in the samples.

The identification of these estrogens was confirmed by the four criteria mentioned in the “Quality controls” section. Nonetheless, it was impossible to confirm the identification of glucocorticoid compounds as no signal was detected for the confirmation transition. Consequently, reported concentrations for these compounds are given only as indications.

$\text{E}1$  was the only compound detected in all samples and was found at concentrations ranging between  $0.19 \pm 0.11$  and  $1.78 \pm 0.39 \text{ ng L}^{-1}$ . COL and  $\text{E}3$  were detected in at least three samples in each campaign.  $17\alpha\text{EE}2$  was detected in one sample at a very low concentration. Among the few comparable data available, Lardy-Fontan et al. reported concentrations in whole water ranging between 0.4 and  $7.71 \text{ ng L}^{-1}$  for  $\text{E}1$  and between 0.4 and  $1.30 \text{ ng L}^{-1}$  for  $17\beta\text{E}2$ , and concentrations ranging between 0.1 and  $0.23 \text{ ng L}^{-1}$  for  $17\alpha\text{EE}2$ , in French surface waters [10]. In the Sambre river, Glineur et al. quantified  $\text{E}1$  at  $0.67 \text{ ng L}^{-1}$  and only detected  $17\beta\text{E}2$ ;  $\text{E}3$  and  $17\alpha\text{EE}2$  were not detected [22]. Even if the occurrence of  $17\alpha\text{EE}2$ ,  $17\beta\text{E}2$ , and  $\text{E}1$  has been extensively studied in the literature [46], most published methods only target the dissolved fraction, and only a few data are available on whole water [37]. Moreover, very little data on the GLUCO, PROG, or ANDRO occurrence are available in the total fraction of inland European waters. T,  $\text{E}1$ ,  $17\beta\text{E}2$ ,  $17\alpha\text{EE}2$ , 17HPT, COR, COL, and DEX were detected in surface waters in Northern Italy at levels ranging between 2 and  $3 \text{ ng L}^{-1}$  for glucocorticoids and  $76 \text{ ng L}^{-1}$  for  $\text{E}1$  [47]. T and AD were detected in drinking water collected in the Pécs in Hungary at 0.013 and  $0.014 \text{ ng L}^{-1}$ , respectively [48]. It is nonetheless difficult to put the data from the present study into perspective with other published works, notably because of the issue of the representativeness of samples [3, 49].

The concentrations reported in this study for WFD-regulated estrogens and COL were lower than the PNEC or AA-EQS for  $17\beta\text{E}2$  and  $\text{E}1$ . Considering this aspect, a good status of the investigated rivers can be derived for these compounds. However, traces of  $17\alpha\text{EE}2$  were detected at a concentration above the recently updated AA-EQS ( $0.017 \text{ ng}$

**Table 3** Measured concentrations of target compounds in Belgian waters

	LQ ( $\text{ng L}^{-1}$ )		Concentration in $\text{ng L}^{-1}$			
			Ourthe River	Meuse River	Lesse River	Sambre River
BET	0.1	December 2020	ND	ND	ND	$0.10 \pm 0.06$
COL	0.5	July 2020	Traces (0.32)	Traces (0.24)	Traces (0.19)	ND
		December 2020	Traces (0.22)	Traces (0.36)	Traces (0.37)	$0.56 \pm 0.24$
COR	0.5	July 2020	Traces (0.21)	ND		
		December 2020	ND	Traces (0.31)	ND	$0.54 \pm 0.31$
$17\alpha\text{EE}2$	0.035	December 2020	ND			Traces (0.024)
$\text{E}3$	0.1	July 2020	$0.15 \pm 0.10$	$0.67 \pm 0.15$	ND	$0.16 \pm 0.11$
		December 2020	Traces (0.07)	$0.51 \pm 0.12$	$0.14 \pm 0.10$	$0.74 \pm 0.17$
$17\alpha\text{E}2$	0.1	December 2020	ND	Traces (0.06)	$0.10 \pm 0.05$	Traces (0.06)
$17\beta\text{E}2$	0.1	July 2020	$0.12 \pm 0.09$	Traces (0.08)	ND	$0.10 \pm 0.07$
		December 2020	ND	Traces (0.08)	ND	$0.12 \pm 0.09$
$\text{E}1$	0.1	July 2020	$0.93 \pm 0.27$	$0.56 \pm 0.16$	$0.35 \pm 0.20$	$0.61 \pm 0.18$
		December 2020	$0.19 \pm 0.11$	$0.70 \pm 0.20$	$0.34 \pm 0.20$	$1.78 \pm 0.39$

ND: compound not detected considering a limit of detection (LD) equivalent to LQ/3

When compounds were detected at levels above the LD and below LQ, concentrations are given as indicative data and mentioned in brackets as “Traces”

$L^{-1}$  since November 2021), which is lower than the LQ validated in this study ( $0.035 \text{ ng L}^{-1}$ ) [36]. Therefore, further investigation to achieve a LQ that can meet the WFD requirement should be implemented. As no PNEC or EQS exist to date for the other compounds, the potential risk to the environment could not be assessed. Soon, closer attention should be brought to LEV and NOR since they are good candidates to be included in the next watchlist.

This study allowed getting information on the occurrence of GLUCO in European rivers. Further monitoring for these compounds is, however, needed to gather more data for their environmental risk assessment.

## Conclusions

For the first time, an isotope dilution two-step SPE-LC-MS/MS method for the characterization of natural and synthetic hormones of four classes (ANDRO, GLUCO, ESTRO, PROG) in the total fraction of surface waters was validated according to current French and European standards (NF T90:210 and XP CEN/TS 16800) in representative matrices and proved to be robust and sensitive.

Regarding existing ecotoxicity data (PNEC), and WFD regulation requirements, the performances of the method were globally considered satisfactory in terms of LQ and accuracy. However, only semi-quantitative data could be obtained for six compounds (DES, Ac CHLOR, Ac CYP, CYP, NOR, and COL), bringing a first level of information on these very few documented compounds in the context of environmental monitoring. An improvement in the method performances might be easily reached through the improvement of the purification step or the systematic use of isotope dilution, for example.

Even if very rarely addressed in published scientific papers, the metrological traceability and the estimation of measurement uncertainty are mandatory to ensure data quality, efficient data interpretation, and comparability of measurement results within laboratories and, more generally, Europe. Therefore, these fundamental metrological aspects were cautiously and thoroughly investigated in the present study. Measurement traceability to the SI units was established using available CRMs. Measurement uncertainties were fully evaluated following the “Guide to the expression of Uncertainty in Measurement” (GUM) under intermediate precision conditions and not under repeatability conditions, as it is too often the case in most published works.

In this study, a detailed methodological validation rarely seen in literature was conducted based on stringent quality controls to analyze these four classes of hormones in surface waters. It also underlined the importance of implementing method validation using representative matrices and the complexity of reaching confident and accurate data at trace levels for such a wide range of target compounds.

The validated method was then successfully applied to environmental monitoring in Belgium. Occurrence of some ESTRO and GLUCO was observed in the sampled Belgian rivers, underlining the need to continue tracking these substances in the aquatic environment to better understand their fate, occurrence, and ecotoxicity, but also to support risk assessment, monitor prioritization, and address the scientific need of characterization for target compounds in the total fraction of surface waters, particularly for non-estrogenic compounds.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00216-023-04698-4>.

**Acknowledgements** The authors would like to thank Dr. Nathalie Guigues from LNE for her support during the writing of the present paper.

**Author contribution** Conceptualization: Elodie Mirmont, Amandine Boeuf, Sophie Lardy-Fontan

Methodology: Elodie Mirmont, Amandine Boeuf, Sophie Lardy-Fontan

Investigation: Elodie Mirmont, Mélissa Charmel

Writing original draft: Elodie Mirmont, Amandine Boeuf, Sophie Lardy-Fontan

Writing review and editing: Elodie Mirmont, Amandine Boeuf, Mélissa Charmel, Béatrice Lalère, Sophie Lardy-Fontan

Visualization: Elodie Mirmont

Supervision: Amandine Boeuf, Sophie Lardy-Fontan

Project administration: Amandine Boeuf, Sophie Lardy-Fontan

Funding acquisition: Amandine Boeuf, Béatrice Lalère, Sophie Lardy-Fontan

**Funding** The ANRT (Association Nationale de la Recherche et de la Technologie) (2017/1131) funded Elodie Mirmont’ PhD work.

**Data availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Code availability** Not applicable.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

## References

1. Wang Y, Zhou J. Endocrine disrupting chemicals in aquatic environments: a potential reason for organism extinction? *Aquat Ecosyst Health Manag*. 2013. <https://doi.org/10.1080/14634988.2013.759073>.
2. World Health Organization, United Nations Environment Programme, Inter-Organization Programme for the Sound Management of Chemicals, Bergman, Åke, Heindel, Jerrold J. et al. State of the science of endocrine disrupting chemicals 2012 : summary for decision-makers. World Health Organization. 2013. <https://apps.who.int/iris/handle/10665/78102>. Accessed 28 Feb 2022.



3. Ojogoro JO, Scrimshaw MD, Sumpter JP. Steroid hormones in the aquatic environment. *Sci Total Environ*. 2021. <https://doi.org/10.1016/j.scitotenv.2021.148306>.
4. Kidd KA, Blanchfield PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, Flick RW. Collapse of a fish population after exposure to a synthetic estrogen. *PNAS*. 2007. <https://doi.org/10.1073/pnas.0609568104>.
5. Runnalls TJ, Beresford N, Losty E, Scott AP, Sumpter JP. Several synthetic progestins with different potencies adversely affect reproduction of fish. *Environ Sci Technol*. 2013. <https://doi.org/10.1021/es3048834>.
6. McNeil PL, Nebot C, Sloman KA. Physiological and behavioral effects of exposure to environmentally relevant concentrations of prednisolone during Zebrafish (*Danio rerio*) embryogenesis. *Environ Sci Technol*. 2016. <https://doi.org/10.1021/acs.est.6b00276>.
7. EU. Commission implementing decision (EU) 2020/1161 of 4 August 2020 establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/EC of the European Parliament and of the Council (notified under document number C(2020) 5205). 2020. [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2020.257.01.0032.01.ENG&toc=OJ:L:2020:257:TOC](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2020.257.01.0032.01.ENG&toc=OJ:L:2020:257:TOC). Accessed 26 Apr 2023
8. Sousa JCG, Ribeiro AR, Barbosa MO, Pereira MFR, Silva AMT. A review on environmental monitoring of water organic pollutants identified by EU guidelines. *J Hazard Mater*. 2018. <https://doi.org/10.1016/j.jhazmat.2017.09.058>.
9. Du B, Fan G, Yu W, Yang S, Zhou J, Luo J. Occurrence and risk assessment of steroid estrogens in environmental water samples: a five-year worldwide perspective. *Environ Pollut*. 2020. <https://doi.org/10.1016/j.envpol.2020.115405>.
10. Lardy-Fontan S, Le Diouren V, Fallot C, Vaslin-Reimann S, Lalere B. Toward the determination of estrogenic compounds in the framework of EU watch list: validation and implementation of a two-step solid phase extraction–liquid phase chromatography coupled to tandem mass spectrometry method. *Accred Qual Assur*. 2018. <https://doi.org/10.1007/s00769-018-1346-4>.
11. Weizel A, Schlüsener MP, Dierkes G, Ternes TA. Occurrence of glucocorticoids, mineralocorticoids, and progestogens in various treated wastewater, rivers, and streams. *Environ Sci Technol*. 2018. <https://doi.org/10.1021/acs.est.7b06147>.
12. Golovko O, Šauer P, Fedorova G, Kroupová HK, Grabic R. Determination of progestogens in surface and waste water using SPE extraction and LC-APCI/APPI-HRPS. *Sci Total Environ*. 2018. <https://doi.org/10.1016/j.scitotenv.2017.10.120>.
13. Goeury K, Vo Duy S, Munoz G, Prévost M, Sauvé S. Analysis of Environmental Protection Agency priority endocrine disruptor hormones and bisphenol A in tap, surface and wastewater by online concentration liquid chromatography tandem mass spectrometry. *J Chromatogr A*. 2019. <https://doi.org/10.1016/j.chroma.2019.01.016>.
14. Zhang K, Zhao Y, Fent K. Occurrence and ecotoxicological effects of free, conjugated, and halogenated steroids including 17 $\alpha$ -hydroxypregnanolone and pregnanediol in swiss wastewater and surface water. *Environ Sci Technol*. 2017. <https://doi.org/10.1021/acs.est.7b01231>.
15. Huysman S, Van Meulebroek L, Vanryckeghem F, Van Langenhove H, Demeestere K, Vanhaecke L. Development and validation of an ultra-high performance liquid chromatographic high resolution Q-Orbitrap mass spectrometric method for the simultaneous determination of steroidal endocrine disrupting compounds in aquatic matrices. *Anal Chim Acta*. 2017. <https://doi.org/10.1016/j.aca.2017.07.001>.
16. Zhang K, Fent K. Determination of two progestin metabolites (17 $\alpha$ -hydroxypregnanolone and pregnanediol) and different classes of steroids (androgens, estrogens, corticosteroids, progestins) in rivers and wastewaters by high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). *Sci Total Environ*. 2018. <https://doi.org/10.1016/j.scitotenv.2017.08.114>.
17. Kase R, Javurkova B, Simon E, Swart K, Buchinger S, Könemann S, Escher BI, Carere M, Dulio V, Ait-Aissa S, Hollert H, Valsecchi S, Polesello S, Behnisch P, di Paolo C, Olbrich D, Sychrova E, Gundlach M, Schlichting R, Leborgne L, Clara M, Scheffknecht C, Marneffe Y, Chalou C, Tusil P, Soldan P, von Danwitz B, Schwaiger J, Palao AM, Bersani F, Perceval O, Kienle C, Vermeirssen E, Hilscherova K, Reifferscheid G, Werner I. Screening and risk management solutions for steroidal estrogens in surface and wastewater. *Trends Anal Chem*. 2018;102:343–58. <https://doi.org/10.1016/j.trac.2018.02.013>.
18. Könemann S, Kase R, Simon E, Swart K, Buchinger S, Schlüsener M, Hollert H, Escher BI, Werner I, Ait-Aïssa S, Vermeirssen E, Dulio V, Valsecchi S, Polesello S, Behnisch P, Javurkova B, Perceval O, Di Paolo C, Olbrich D, Sychrova E, Schlichting R, Leborgne L, Clara M, Scheffknecht C, Marneffe Y, Chalou C, Tušil P, Soldan P, von Danwitz B, Schwaiger J, San Martín Becares MI, Bersani F, Hilscherová K, Reifferscheid G, Ternes T, Carere M. Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive. *Trends Anal Chem*. 2018. <https://doi.org/10.1016/j.trac.2018.02.008>.
19. Lai KM, Johnson KL, Scrimshaw MD, Lester JN. Binding of waterborne steroid estrogens to solid phases in river and estuarine systems. *Environ Sci Technol*. 2000. <https://doi.org/10.1021/es9912729>.
20. Martínez-Hernández V, Meffe R, Herrera S, Arranz E, de Bustamante I. Sorption/desorption of non-hydrophobic and ionisable pharmaceutical and personal care products from reclaimed water onto/from a natural sediment. *Sci Total Environ*. 2014. <https://doi.org/10.1016/j.scitotenv.2013.11.036>.
21. EU. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. 2000. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0060>. Accessed 26 Apr 2023
22. Glineur A, Barbera B, Nott K, Carbone P, Ronkart S, Lognay G, Tyteca E. Trace analysis of estrogenic compounds in surface and ground water by ultra high performance liquid chromatography-tandem mass spectrometry as pyridine-3-sulfonyl derivatives. *J Chromatogr A*. 2018. <https://doi.org/10.1016/j.chroma.2017.12.042>.
23. Ten KA. years of mixing cocktails: a review of combination effects of endocrine-disrupting chemicals. *Environ Health Perspect*. 2007. <https://doi.org/10.1289/ehp.9357>.
24. Mirmont E, Bœuf A, Charmel M, Lalère B, Laprêvotte O, Lardy-Fontan S. Overcoming matrix effects in quantitative liquid chromatography–mass spectrometry analysis of steroid hormones in surface waters. *Rapid Commun Mass Spectrom*. 2022. <https://doi.org/10.1002/rcm.9154>.
25. AFNOR. XP CEN/TS 16800. In: Afnor EDITIONS. 2021. <https://www.boutique.afnor.org/fr-fr/norme/xp-cen-ts-16800/lignes-directrices-pour-la-validation-des-methodes-danalyse-physicochimique/fa201876/238965>. Accessed 12 May 2022.
26. AFNOR. NF T90-210. In: Afnor EDITIONS. 2018. <https://www.boutique.afnor.org/fr-fr/norme/nf-t90210/qualite-de-leau-protocole-devaluation-initiale-des-performances-dune-method/fa190833/81632>. Accessed 12 May 2022.
27. JCGM. Guides to the expression of uncertainty in measurement. 2020. <https://www.iso.org/sites/JCGM/GUM-introduction.htm>. Accessed 28 Feb 2022.
28. Elordui-Zapatarietxe S, Fettig I, Philipp R, Gantois F, Lalère B, Swart C, Petrov P, Goenaga-Infante H, Vanermen G, Boom G, Emteborg H. Novel concepts for preparation of reference materials as whole water samples for priority substances at nanogram-per-liter level using model suspended particulate matter and

- humic acids. *Anal Bioanal Chem.* 2015. <https://doi.org/10.1007/s00216-014-8349-8>.
29. Mirmont E, Bœuf A, Charmel M, Vaslin-Reimann S, Lalère B, Laprévotte O, Lardy-Fontan S. Development and implementation of an analytical procedure for the quantification of natural and synthetic steroid hormones in whole surface waters. *J Chromatogr B.* 2021;1175:122732. <https://doi.org/10.1016/j.jchromb.2021.122732>.
30. Database of higher-order reference materials, measurement methods/procedures and services. <https://www.bipm.org/jctlm/news.do>. Accessed 28 Feb 2022.
31. Duxbury K, Owen L, Gillingwater S, Keevil B. Naturally occurring isotopes of an analyte can interfere with doubly deuterated internal standard measurement. *Ann Clin Biochem.* 2008. <https://doi.org/10.1258/acb.2007.007137>.
32. Davison AS, Milan AM, Dutton JJ. Potential problems with using deuterated internal standards for liquid chromatography-tandem mass spectrometry. *Ann Clin Biochem.* 2013. <https://doi.org/10.1177/0004563213478938>.
33. ISO. ISO 21253-1:2019. In: ISO. <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/02/70256.html>. Accessed 28 Feb 2022.
34. AFNOR. FD T90-230. In: Afnor EDITIONS. <https://www.boutique.afnor.org/fr-fr/norme/fd-t90230/qualite-de-leau-caracterisation-des-methodes-danalyses-guide-pour-la-select/fa182577/46180>. Accessed 28 Feb 2022.
35. JCGM. International vocabulary of metrology. 2012. <https://www.iso.org/sites/JCGM/VIM-introduction.htm>. Accessed 28 Feb 2022.
36. EU, Scientific Opinion on “Draft Environmental Quality Standards for Priority Substances under the WFD”-17-Alpha-Ethynylestradiol (EE2), Beta-Estradiol (E2) and Estrone (E1). 2022. [https://ec.europa.eu/health/publications/scientific-opinion-draft-environmental-quality-standards-priority-substances-under-wfd-17-alpha\\_en](https://ec.europa.eu/health/publications/scientific-opinion-draft-environmental-quality-standards-priority-substances-under-wfd-17-alpha_en). Accessed 21 Jun 2022.
37. Glineur A, Nott K, Carbonnelle P, Ronkart S, Purcaro G. Development and validation of a method for determining estrogenic compounds in surface water at the ultra-trace level required by the EU Water Framework Directive Watch List. *J Chromatogr A.* 2020. <https://doi.org/10.1016/j.chroma.2020.461242>.
38. Zhang J-N, Chen J, Yang L, Zhang M, Yao L, Liu Y-S, Zhao J-L, Zhang B, Ying G-G. Occurrence and fate of androgens, progestogens and glucocorticoids in two swine farms with integrated wastewater treatment systems. *Water Res.* 2021. <https://doi.org/10.1016/j.watres.2021.116836>.
39. Shen X, Chang H, Sun D, Wang L, Wu F. Trace analysis of 61 natural and synthetic progestins in river water and sewage effluents by ultra-high performance liquid chromatography-tandem mass spectrometry. *Water Res.* 2018. <https://doi.org/10.1016/j.watres.2018.01.030>.
40. Shen X, Chang H, Sun Y, Wan Y. Determination and occurrence of natural and synthetic glucocorticoids in surface waters. *Environ Int.* 2020. <https://doi.org/10.1016/j.envint.2019.105278>.
41. Leusch F, Neale P, Arnal C, Aneck-Hahn N, Balaguer P, Bruchet A, Escher B, Esperanza M, Grimaldi M, Leroy G, Scheurer M, Schlichting R, Schriks M, Hebert A. Analysis of endocrine activity in drinking water, surface water and treated wastewater from six countries. *Water Res.* 2018. <https://doi.org/10.1016/j.watres.2018.03.056>.
42. Goeury K, Vo Duy S, Munoz G, Prévost M, Sauvé S. Assessment of automated off-line solid-phase extraction LC-MS/MS to monitor EPA priority endocrine disruptors in tap water, surface water, and wastewater. *Talanta.* 2022. <https://doi.org/10.1016/j.talanta.2022.123216>.
43. Tavazzi S, Mariani G, Comero S, Ricci M, Paracchini B, Skejo H, Gawlik B. Water Framework Directive Watch List Method Analytical method for determination of compounds selected for the first Surface Water Watch List. Publications Office of the European Union. 2016.
44. EU. Commission Directive 2009/90/EC of 31 July 2009 laying down, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, technical specifications for chemical analysis and monitoring of water status. 2009. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32009L0090>. Accessed 26 Apr 2023
45. ISO, ISO 11352:2012. In: ISO. <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/03/50399.html>. Accessed 28 Feb 2022.
46. Sousa JCG, Ribeiro AR, Barbosa MO, Ribeiro C, Tiritan ME, Pereira MFR, Silva AMT. Monitoring of the 17 EU Watch List contaminants of emerging concern in the Ave and the Sousa Rivers. *Sci Total Environ.* 2019. <https://doi.org/10.1016/j.scitotenv.2018.08.309>.
47. Merlo F, Speltini A, Maraschi F, Sturini M, Profumo A. HPLC-MS/MS multiclass determination of steroid hormones in environmental waters after preconcentration on the carbonaceous sorbent HA-C@silica. *Arab J Chem.* 2020. <https://doi.org/10.1016/j.arabjc.2019.10.009>.
48. Molnár S, Kulcsár G, Perjési P. Determination of steroid hormones in water samples by liquid chromatography electrospray ionization mass spectrometry using parallel reaction monitoring. *Microchem J.* 2022. <https://doi.org/10.1016/j.microc.2021.107105>.
49. Petrovic M. Methodological challenges of multi-residue analysis of pharmaceuticals in environmental samples. *Trends Environ Anal Chem.* 2014. <https://doi.org/10.1016/j.teac.2013.11.004>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.